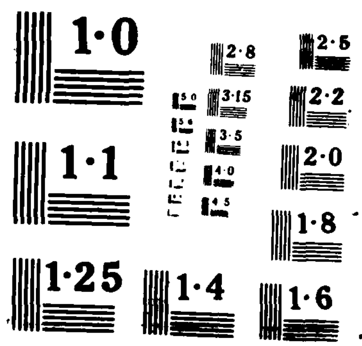
AD-A176 307 USING INFLUENCE DIAGRAMS TO SOLVE THE CALIBRATION 1/1
PROBLEM(U) CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH
CENTER R E BARLOW ET AL. AUG 86 ORC-86-13
UNCLASSIFIED AFOSR-81-0122 F/G 12/2 NL


1·0
2·8
2·5
3·15
2·2
3·5
1·1
2·0
4·0
4·5
1·8
1·25
1·4
1·6

OPERATIONS RESEARCH CENTER

USING INFLUENCE DIAGRAMS
TO SOLVE THE CALIBRATION PROBLEM*

by

Richard E. Barlow[1]
R. W. Mensing[2]
N. C. Smiriga[3]

UNIVERSITY OF CALIFORNIA
BERKELEY

11

USING INFLUENCE DIAGRAMS TO SOLVE THE CALIBRATION PROBLEM*

by

Richard E. Barlow[1]

R. W. Mensing[2]

N. G. Smiriga[3]

ORC 86-13 August 1986

*This research was supported by the Air Force Office of Scientific Research (AFOSR), USAF under Grant AFOSR-81-0122 and the U. S. Army Research Office under Contract DAAG29-85-K-0208 with the University of California and the Department of Statistics, Warwick University, England.

[1]Department of Industrial Engineering & Operations Research, Dept. of Statistics, University of California, Berkeley, California 94720.

[2]The Statistics Department, University of Warwick, Coventry CV4 7AL.

[3]Lawrence Livermore National Laboratory, Livermore, CA 94550.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

ADA 176387

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ORC 86-13	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Operations Research Center 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State and ZIP Code) 3115 Etcheverry Hall University of California Berkeley, CA 94720	7b. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-81-0122
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB, DC	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 6.1102F PROJECT NO.: 2304 TASK NO.: WORK UNIT NO.:
11. TITLE (Include Security Classification) Using Influence Diagrams to Solve the Calibration Problem	
12. PERSONAL AUTHOR(S) Richard E. Barlow, R.W. Mensing, N.G. Smiriga	
13a. TYPE OF REPORT Technical 13b. TIME COVERED FROM ______ TO ______	14. DATE OF REPORT (Yr., Mo., Day) August 1986 15. PAGE COUNT 36
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
19. ABSTRACT (Continue on reverse if necessary and identify by block number) See Report	
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Major Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code) 202-767-5027 22c. OFFICE SYMBOL AFOSR/NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

ACKNOWLEDGEMENTS

We would like to acknowledge Tony O'Hagan for helpful suggestions and for providing facilities for much of this research in connection with the Bayesian Study year at Warwick University. We would also like to thank Morie De Groot and Bill Jewell for helpful comments and advice as well as Sung Chul Kim for carefully checking calculations and computer programs. Ross Shachter taught us how to use influence diagrams.

USING INFLUENCE DIAGRAMS
TO SOLVE THE CALIBRATION PROBLEM

1. INTRODUCTION

A measuring instrument measures a unit and records an observation y. The "true" measurement, x, of the unit is to be inferred from y. If $p(y|x)$ is the likelihood of y given x and x has prior $p(x)$, then by Bayes' Theorem

$$p(x|y) \propto p(y|x)p(x).$$

Let x_0 and σ_0^2 be the mean and variance of $p(x)$. We will assess the likelihood, $p(y|x)$, using a linear regression model

$$y = \alpha + \beta(x - x^*) + \epsilon \qquad (1.1)$$

where x^* is specified and a priori $(\alpha,\beta) \perp x \perp \epsilon$ and ϵ given x^* is $N(0,\sigma^2)$ with σ specified. (These assumptions could of course be relaxed; e.g. σ^2 unknown, ϵ dependent on x, etc. However our assumptions are convenient and sufficiently general to provide conclusions of general interest.)

The "center", x^*, of the likelihood model and the prior for x are intertwined. The natural choice for x^* is the mean of the prior for x, namely $x^* = x_0$. This is reasonable since our attention is focused on calculating $p(x|y)$. The line, with $x^* = x_o$, is $y = \alpha + \beta(x - x_o)$ where α and β are unknown and of course y cannot be observed without error. See Figure 1.1. Of course the prior for (α, β) depends on $x^* = x_o$ and we may write $p(\alpha, \beta) = p(\alpha|\beta, x_o)\ p(\beta)$ since, in general, only α depends on x_o.

Figure 1.2 is an influence diagram describing the logical and statistical dependencies between unknown quantities, decision alternatives and values (losses or utilities). The decision may be an estimate for x given y. If the worth or loss is

$$w(d,x) = (d - x)^2$$

then the optimal decision will be the posterior mean for x given y.

The Calibration Experiment

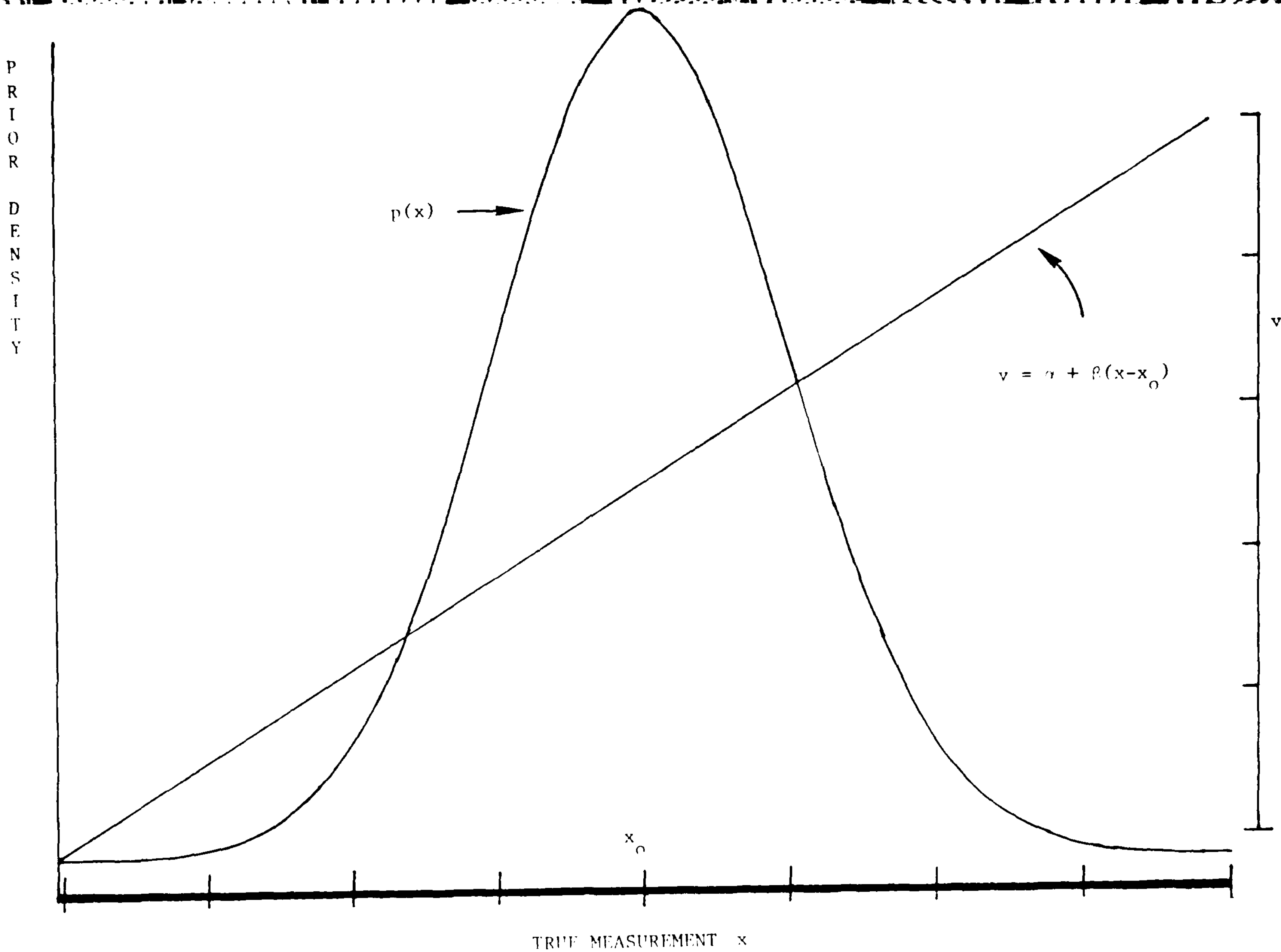
The purpose of the calibration experiment is to learn about (α, β) so that given a future observation y we can reduce our uncertainty about a future "true" measurement x. To calibrate our measuring instrument, we record n measurements

$$y = (y_1, y_2, \ldots, y_n)$$

on n units all of whose "true" measurements,

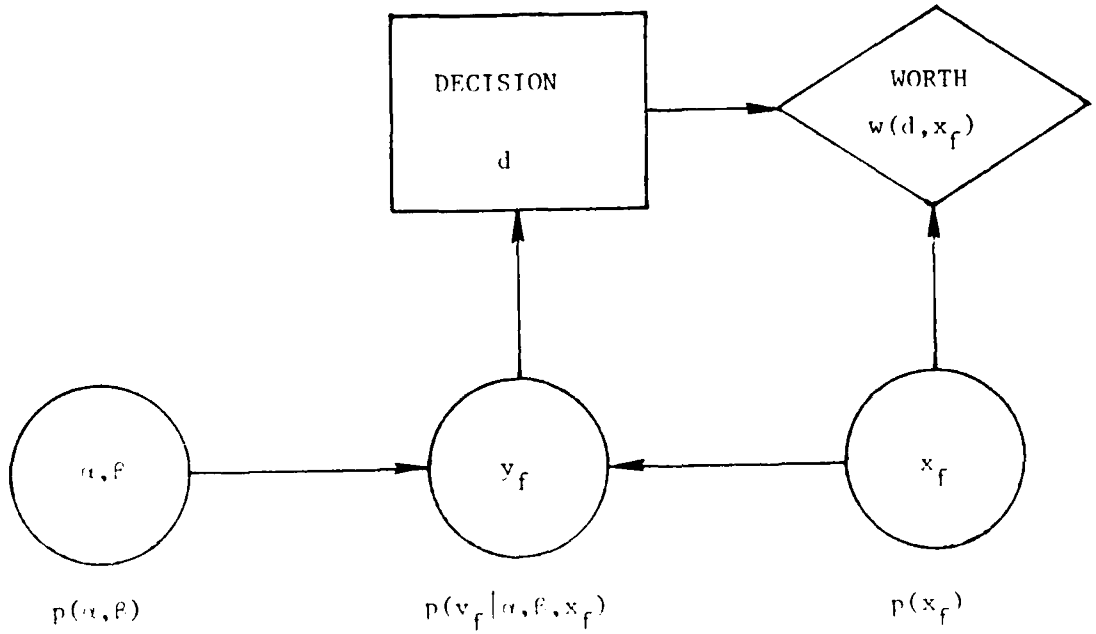
$$x = (x_1, x_2, \ldots, x_n)$$

are specified beforehand. Based on our prior, $p(x)$, for a future x (call it x_f) and our regression model (1.1), our problem is to determine $x = (x_1, x_2, \ldots, x_n)$ (subject to feasibility constraints) so as to minimize some overall loss function. x is called the experimental design for the calibration experiment.



THE CALIBRATION LINE

Figure 1.1



INFLUENCE DIAGRAM

FOR THE CALIBRATION PROBLEM

Figure 1.2

The following assumptions will be made relative to the calibration experiment.

Assumption 1. The future "true value", x_f, is independent of (α,β), x and y. The future observation, y_f, is independent of (x,y) given (α,β).

Assumption 2. The worth function $w(d,x_f)$ is a loss function and depends only on d (the decision regarding x_f taken at the time we observe y_f) and the "true value" x_f. For example, we are ignoring the cost of performing the experiment.

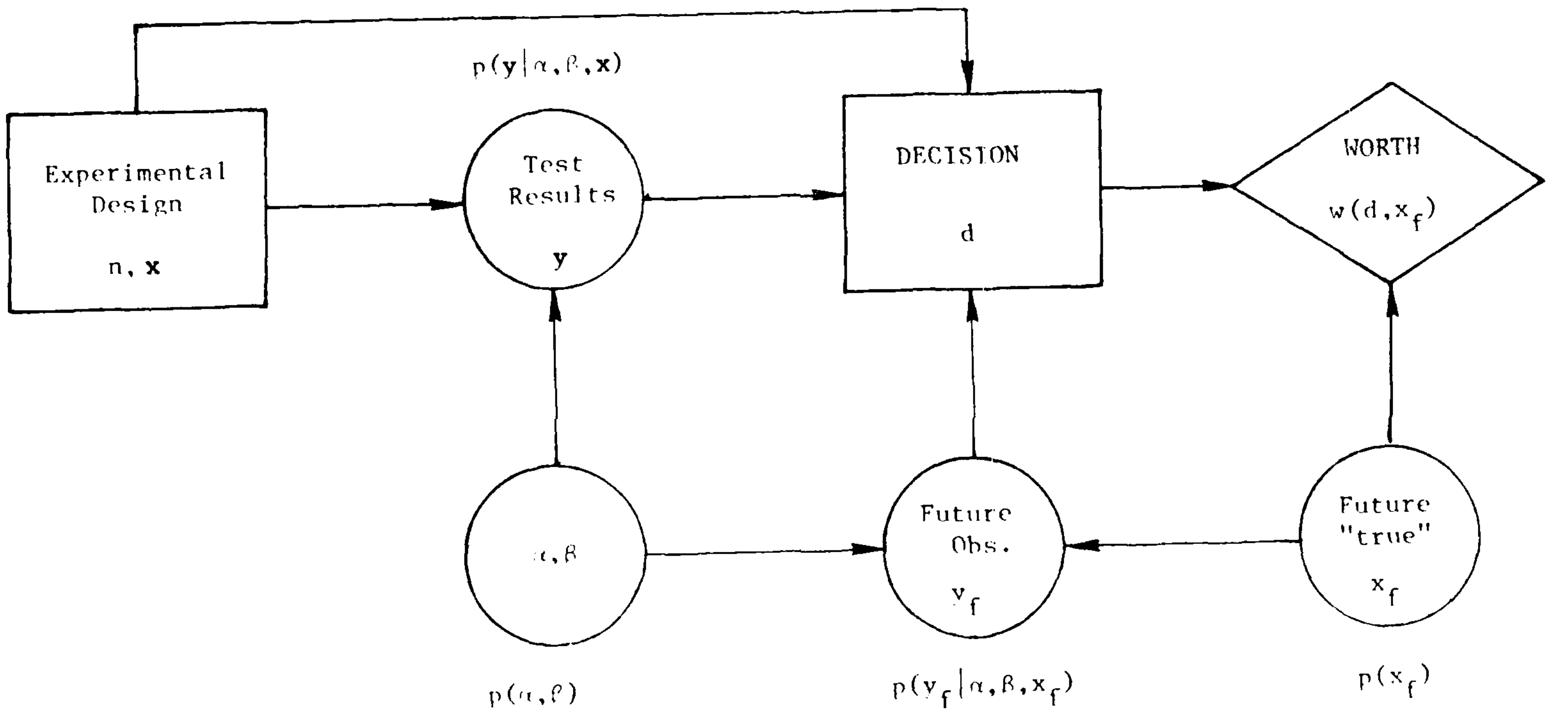
Assumption 3. The feasible region, R, for the experimental design, x, is bounded. That is, infinite x_i values are not allowed in practice. We seek an optimal experimental design subject to $x \in R$.

Figure 1.3 is an influence diagram describing the logical and statistical dependencies between the unknown quantities and decision variables in our problem. Figure 1.4 shows the influence diagram after (α,β) have been eliminated by computing the posterior distribution, $p(\alpha,\beta|x,y)$ and then calculating

$$p(y_f|\ x,y,x_f) = \iint p(y_f|\ \alpha,\beta,x_f)\ p(\alpha,\beta|\ x,y)\ d\alpha\ d\beta.$$

[Influence diagram operations are discussed in Shachter (1986) and Barlow (1987).]

If we take squared error loss, $(d - x_f)^2$, as our worth function where x_f is the future "true" measurement of a unit, then d is our estimate of x_f after we observe a future y_f. Note that at the time of decison when we estimate x_f, we know x, y and y_f. Since we do not know x_f at this time, $p(x_f|y_f,y,x)$ must be found via Bayes' Theorem and

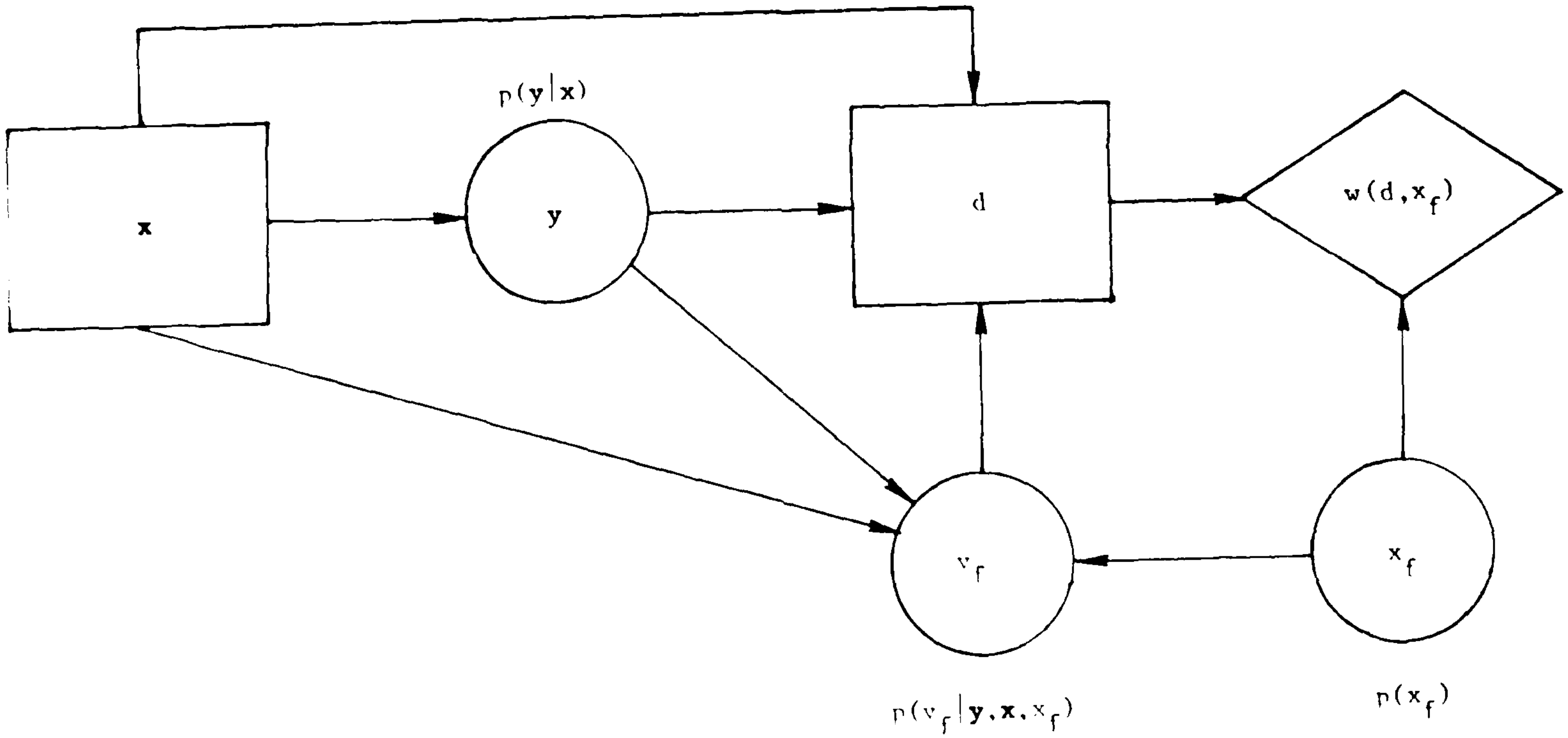


$y_f = \alpha + \beta\,(x_f - x_o) + \varepsilon$

$\varepsilon \sim N(0,\sigma^2)$

THE CALIBRATION PROBLEM

Figure 1.3



THE CALIBRATION PROBLEM

(After Elimination of ([illegible]))

Figure 1.[illegible]

$$\int_{-o}^{\infty} (d - x_f)^2 \; p(x_f | y_f, y, x) \; dx_f$$

calculated. The minimizer is the posterior mean

$$d = E(x_f | y_f, y, x)$$

so that after observing y_f, our worth function is

$$Var(x_f | y_f, y, x).$$

At the design stage, we of course do not know y_f or the test results y. Hence, using the method of Bayesian decision analysis we must minimize

$$E_{y|x} E_{y_f|y,x} \operatorname*{Min}_{d} E_{x_f|y_f,y,x}[w(d,x_f)|y_f,y,x] \overset{DEF}{=} W(x) \qquad (1.2)$$

with respect to $x = (x_1, x_2, \ldots, x_n)$. W(x) is the final expected worth function with respect to the experimental design x.

For a more detailed discussion of this problem and references to other approaches see Chapter 10 of Aitchison and Dunsmore (1980). Hoadley (1970) discusses the calibration inference problem in some detail and points out the difficulties with the maximum likelihood estimator for x_f given an observation y_f and data $\{(x_i, y_i), \; i = 1, 2, \ldots, n\}$ from a calibration experiment. Brown (1982) and Brown and Sundberg (1985) extend Hoadley's results using a multivariate formulation. However they do not consider the problem of optimal Bayesian experimental design. The definitive reference for Bayesian design for linear regression is Chaloner (1984). The objective of this paper is to discuss the calibration experimental design problem and results for special cases.

Summary of Results

Based on the likelihood it is shown that the experimental design may be summarized by n, $\bar{x} - x_o = \sum_i^n (x_i - x_o)/n$ and $v_x^2 = \sum_i^n (x_i - x_o)^2/n$ where

$$|\bar{x} - x_o| \leq v_x.$$

If β is known, W(x) depends only on n and the optimal design corresponds to taking n as large as possible. The values of x are immaterial. If α is known, W(x), depends only on $\sum_i^n (x_i - x_o)^2$ and is decreasing in v_x for fixed n.

If both α and β are unknown, the optimal design can be found by performing a three dimensional search over $(n, \bar{x}, v_x)$. W(x) can be evaluated numerically by using three nested subroutines when the prior for (α,β) is bivariate normal and $w(d,x_f) = (d - x_f)^2$. For this case and x_f $N(x_o,\sigma_o^2)$, we can explicitly calculate $W(x|\sigma_b = 0)$. Also for this case, $W(x|x_1 = x_2 = \ldots = x_n = x_o)$ can be numerically calculated using two nested subroutines.

2. WORTH OF INFORMATION GAINED

Suppose we perform the calibration experiment x. Then

$$\operatorname*{Min}_d E_{x_f}[w(d,x_f)] - W(x) \tag{2.1}$$

is a measure of the expected reduction in our uncertainty about x_f (when $w(d,x_f) = (d - x_f)^2$) as a result of performing the calibration experiment. Lemma 2.1 shows that this difference is ≥ 0. This is the familiar expected information inequality in our notation [Raiffa and Schlaifer (1961)]. It gives us easily computed upper bounds on W(x) as a result of performing the calibration experiment. This is useful in checking computer calculations. From Figure 1.3 we see that at the time of decision (e.g. estimating x_f) we know x, y, and y_f. It is intuitively clear that when $w(d,x_f)$ is a loss function the final expected value will be greater the less information we have at the time of decision.

The following results stated as lemmas will be used in the next section.

Lemma 2.1. If the range of possible decisions, d, does not depend on x or y then

$$E_{y|x} E_{y_f|y,x} \underset{d}{\text{Min}} E_{x_f|y_f,y,x}[w(d,x_f)|y_f,y,x]$$

$$\leq E_{y_f} \underset{d}{\text{Min}} E_{x_f|y_f}[w(d,x_f)|y_f]$$

$$\leq \underset{d}{\text{Min}} E_{x_f}[w(d,x_f)] \quad .$$

Proof. We will prove the first inequality. Let

$$\underset{d}{\text{Min}} E_{x_f|y_f}[w(d,x_f)|y_f] = E_{x_f|y_f}\{w[d_o(y_f),x_f]|y_f\}$$

so that

$$E_{y|x} E_{y_f|y,x} \underset{d}{\text{Min}} E_{x_f|y_f,y,x}\{[w(d,x_f) - w[d_o(y_f),x_f]]|y_f,y,x\} \leq 0.$$

We need only show

$$E_{y|x} E_{y_f|y,x} E_{x_f|y_f,y,x}\{w[d_o(y_f),x_f]|y_f,y,x\}$$
$$= E_{y_f} E_{x_f|y_f}\{[w[d_o(y_f),x_f]|y_f]\}.$$

From Bayes' theorem and the fact that x_f is independent of (x,y) we have

$$p(x_f,y_f|x,y)p(y|x) = [p(y|x_f,y_f,x)p(x_f,y_f)/p(y|x)]p(y|x)$$
$$= p(y|y_f,x)p(x_f,y_f).$$

The result follows by an interchange in the order of integration.

The second inequality follows in a similar way. QED

Remark. $E_{y_f} \operatorname*{Min}_d E_{x_f|y_f} [w(d,x_f)|y_f]$ corresponds to not performing the calibration experiment (i.e., $n = 0$). When $w(d,x_f) = (d - x_f)^2$ the above inequalities become

$$E_{y|x} E_{y_f|y,x} \operatorname{Var}(x_f|y_f,y,x) \le E_{y_f} \operatorname{Var}(x_f|y_f) \le \operatorname{Var}(x_f) \quad .$$

It follows from Lemma 2.1 that the expected worth function can only decrease if we perform additional calibration experiments. We use this fact later. (This would not be true if $w(\cdot,\cdot)$ depended on (x,y).)

Lemma 2.2. Under the assumptions of Lemma 2.1 and $w(d,x_f)$ a loss function

$$W(x_1,\ldots,x_n) \ge W(x_1,\ldots,x_n,x_{n+1})$$

where the first n coordinates are the same on both sides of the inequality.

3. LIKELIHOOD AND THE OPTIMAL EXPERIMENTAL DESIGN

Under the assumption that observation errors, $\{\epsilon_i |\ i = 1,2,\ldots,n\}$ are independent $N(0,\sigma^2)$, but without specifying prior distributions, we can determine some of the structure of the optimal experimental design. This can be done using the sufficient statistics for (α,β) corresponding to our likelihood model. As noted before, the purpose of the calibration experiment is to learn about (α,β). The likelihood for (α,β) given the data is

$$L(\alpha,\beta|\text{Data},x_o) \;\alpha\; \exp\{-\sum_i^n[y_i - \alpha - \beta(x_i - x_o)]^2/2\sigma^2\}. \qquad (3.1)$$

A priori assume $\alpha \perp \beta \perp \epsilon$ and let $E(\alpha) = a$, $E(\beta) = b$, $Var(\alpha) = \sigma_a{}^2$, and $Var(\beta) = \sigma_b{}^2$. Define

$$e_i = y_i - a - b(x_i - x_o)$$

and rewrite

$$\begin{aligned} y_i - \alpha - \beta(x_i-x_o) &= [y_i - a - b(x_i-x_o)] - (\alpha-a) - (\beta-b)(x_i-x_o) \\ &= e_i - (\alpha - a) - (\beta - b)(x_i - x_o) \end{aligned}$$

so that

$$L(\alpha,\beta|\text{Data},x_o)$$

$$\alpha \quad \exp\{-[n(\alpha - a)^2 + (\beta - b)^2 \sum_i^n(x_i - x_o)^2 - 2\sum_i^n e_i[(\alpha - a) + (\beta - b)(x_i - x_o)] + 2(\alpha - a)(\beta - b)\sum_i^n(x_i - x_o)]/2\sigma^2\}. \qquad (3.2)$$

Clearly n, $\sum_1^n(x_i - x_o)$, $\sum_1^n(x_i - x_o)^2$, $z_1 = \sum_1^n e_i$ and $z_2 = \sum_1^n e_i(x_i - x_o)$ are sufficient statistics for (α,β) since x_o, a, b and σ are specified. It follows that the posterior density for (α,β) also depends on the data only through n, $\sum_1^n(x_i - x_o)$, $\sum_1^n(x_i - x_o)^2$, z_1, and z_2.

<u>Theorem 3.1</u>. $W(x)$ depends on x only through n, $\bar{x} - x_o = \sum_i^n(x_i - x_o)/n$ and

$$v_x{}^2 = \sum_1^n(x_i - x_o)^2/n.$$

N. B. This is true for all worth functions $w(d,x_f)$ and priors on (α,β) and x_f. The worth function can also depend on n, $\bar{x} - x_o$ and v_x in this case.

Proof. The purpose of the calibration experiment is to learn about (α,β). Since n, $\bar{x} - x_o$, v_x, z_1 and z_2 are sufficient statistics for (α,β), the test results, y, may be summarized by z_1 and z_2. Hence from (1.2) we need only show that the joint distribution of (z_1,z_2) depends on x only through n, $\bar{x} - x_o$ and v_x.

It is easy to show that (z_1,z_2) given (α,β) is bivariate normal where z_1 given (α,β) is

$$N[n(\alpha - a) + (\beta - b) \sum_1^n(x_1 - x_o),\ n\sigma^2]$$

and z_2 given (α,β) is

$$N[(\alpha - a) \sum_1^n(x_i - x_o) + (\beta - b) \sum_1^n(x_i - x_o)^2,\ \sigma^2 \sum_1^n(x_i - x_0)^2]$$

while

$$\mathrm{Cov}(z_1,\ z_2|\alpha,\beta) = \sigma^2 \sum_1^n(x_i - x_o). \qquad \text{QED}$$

Corollary 3.2. If β is known, i.e. $\sigma_b = 0$, then W(x) depends on x only through n. The "levels" $(x_1,x_2,\dots,x_n)$ are immaterial and we might just as well take

$$x_1 = x_2 = \dots = x_n = x_o$$

or any other values that we like.

Proof. If we are certain that $\beta = b$; i.e. $\sigma_b = 0$, then (3.2) becomes

$$L(\alpha|\mathrm{Data},x_o) \ \alpha\ \exp\{-[n(\alpha - a)^2 - 2\sum_1^n e_i(\alpha - a)]/2\sigma^2\}.$$

Hence n and $z_1 = \sum_1^n e_i = \sum_1^n[y_i - a - b(x_i - x_o)]$ are sufficient for α.

Since z_1 given $(\alpha,\beta=b)$ is

$$N[n(\alpha - a),\ n\sigma^2]$$

it follows that W(x) depends on x only through n. QED

Corollary 3.3. If α is known, i.e. $\sigma_a = 0$, then $W(x)$ depends on x only through $\sum_i^n (x_i - x_o)^2$. Furthermore, for fixed n, $W(x)$ is decreasing in v_x. In this case, $W(x)$ is minimized for those x belonging to R for which v_x is maximum.

Proof. If $\sigma_a = 0$, then (3.2) becomes

$$L(\beta|\text{Data}, x_o) \;\alpha\; \exp\{-[(\beta-b)^2 \sum_1^n (x_i-x_o)^2 - 2(\beta-b)\sum_1^n e_i(x_i-x_o)]/2\sigma^2\}.$$

Hence $\sum_1^n (x_i - x_o)^2$ and $z_2 = \sum_1^n e_i(x_i - x_o)$ are sufficient for β. Since z_2 given $(\alpha = a, \beta)$ is

$$N[(\beta - b)\sum_1^n (x_i - x_o)^2, \sigma^2 \sum_1^n (x_i - x_o)^2]$$

it follows that when $\alpha = a$ is known, $W(x)$ depends on x only through $\sum_1^n (x_i - x_o)^2$.

Suppose $\sum_1^n (x_i - x_o)^2 < \sum_1^n (x_i' - x_o)^2$. Clearly we can find x_{n+1} such that

$$\begin{aligned}\sum_1^n (x_i' - x_o)^2 &= \sum_1^n (x_i - x_o)^2 + (x_{n+1} - x_o)^2 \\ &= \sum_1^{n+1} (x_i - x_o)^2.\end{aligned}$$

From Lemma 2.2 in section 2 we have

$$W(x_1,\ldots,x_n) \geq W(x_1,\ldots,x_n,x_{n+1}) .$$

Hence $W(x)$ is decreasing in $\sum_1^n (x_i - x_o)^2$ for fixed n. QED

Determining the Structure of the Optimal Experimental Design

Since

$$\sum_{i}^{n}(x_i - \bar{x})^2/n \geq 0$$

it follows that

$$\sum_{1}^{n}(x_i - x_o + x_o - \bar{x})^2/n = \sum_{1}^{n}(x_i - x_o)^2/n - (\bar{x}_i - x_o)^2 \geq 0$$

and $$|\bar{x} - x_o| \leq v_x.$$

Consequently, the minimization problem with respect to x can be transformed to a minimization problem with respect to only three variables, namely n and

$$|\bar{x} - x_o| \leq v_x.$$

Since $\bar{x} - x_o$ and v_x are symmetric functions of an experimental design x, it follows that, for fixed n, any permutation of the coordinates of an experimental design solution is also a solution (if allowed by the feasibility constraints). Figure 3.1 shows the nature of the possible (x_1, x_2) solutions for v_x fixed and n = 2. The darkened arcs on the circumference show the possible designs for a fixed v_x (up to permutations of coordinates). For fixed v_x, possible solutions are traced out by the intersection of the line $\bar{x} - x_o = c$ with the circumference of the circle $x_1^2 + x_2^2 = v_x^2$ as c varies from $-v_x$ to v_x.

The optimal experimental design, x, can, in theory, be found through a three dimensional search over the feasible region R. One strategy would be to fix n and, using a computer calculate a three dimensional plot of

$$W(x) = E_{y|x} E_{y_f|y,x} \min_{d} E_{x_f|y_f,y,x}[w(d,x_f)|y_f,y,x]$$

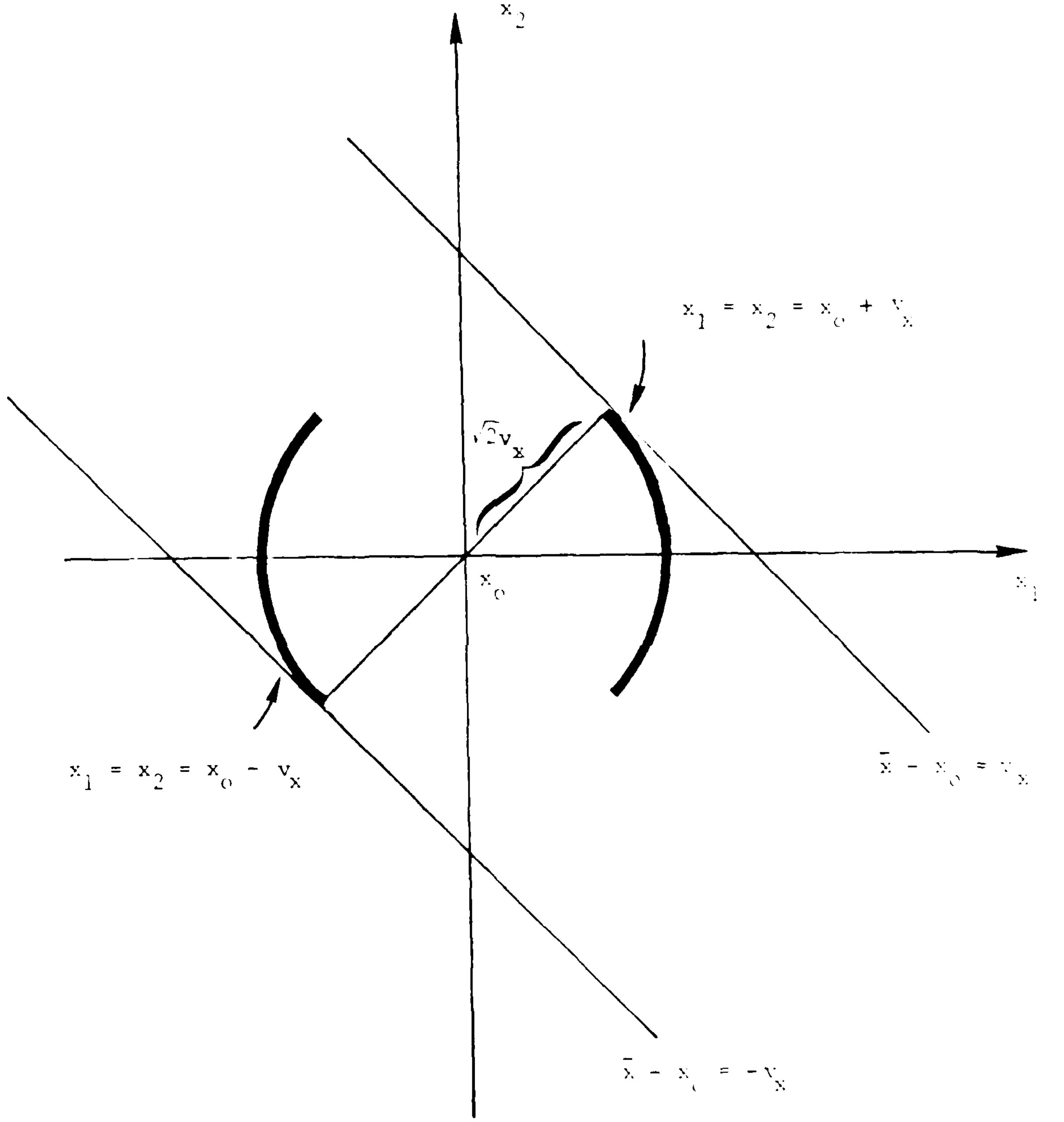


Figure 3.1

n = 2

versus $\bar{x} - x_o$ and v_x. Figure 3.2 illustrates the 3 dimensional plot for a fixed n. The plot shows the surface of W(x) as a function of $|\bar{x} - x_o| \leq v_x$.

The Case $x_1 = x_2 = \ldots = x_n = x_o$

Suppose we are uncertain about both α and β. From (3.2) we see that if $x_1 = x_2 = \ldots = x_n = x_o$, then

$$L(\alpha,\beta \mid \text{Data}) \propto \exp\{-[n(\alpha - a)^2 - 2\sum_1^n e_i(\alpha - a)]/2\sigma^2\}$$

so that in this case the data provide no direct information about β. If in addition, the prior for (α,β) satisfies

$$p(\alpha,\beta|x_o) \propto p(\alpha|x_o)\ p(\beta)$$

i.e. α and β are a priori independent given x_o, then

$$p(\alpha,\beta|\text{Data},x_o) \propto L(\alpha|\text{Data},x_o)\ p(\alpha|x_o)\ p(\beta)$$

and the posterior marginal for β is the same as the prior marginal for β. Intuitively, if β is unknown, the experimental design

$$x_1 = x_2 = \ldots = x_n = x_o$$

is a local maximum for the final expected value since values of x_i near x_o will provide information about β and hence tend to reduce the final expected value.

The Case $w(d,x_f) = (d - x_f)^2$

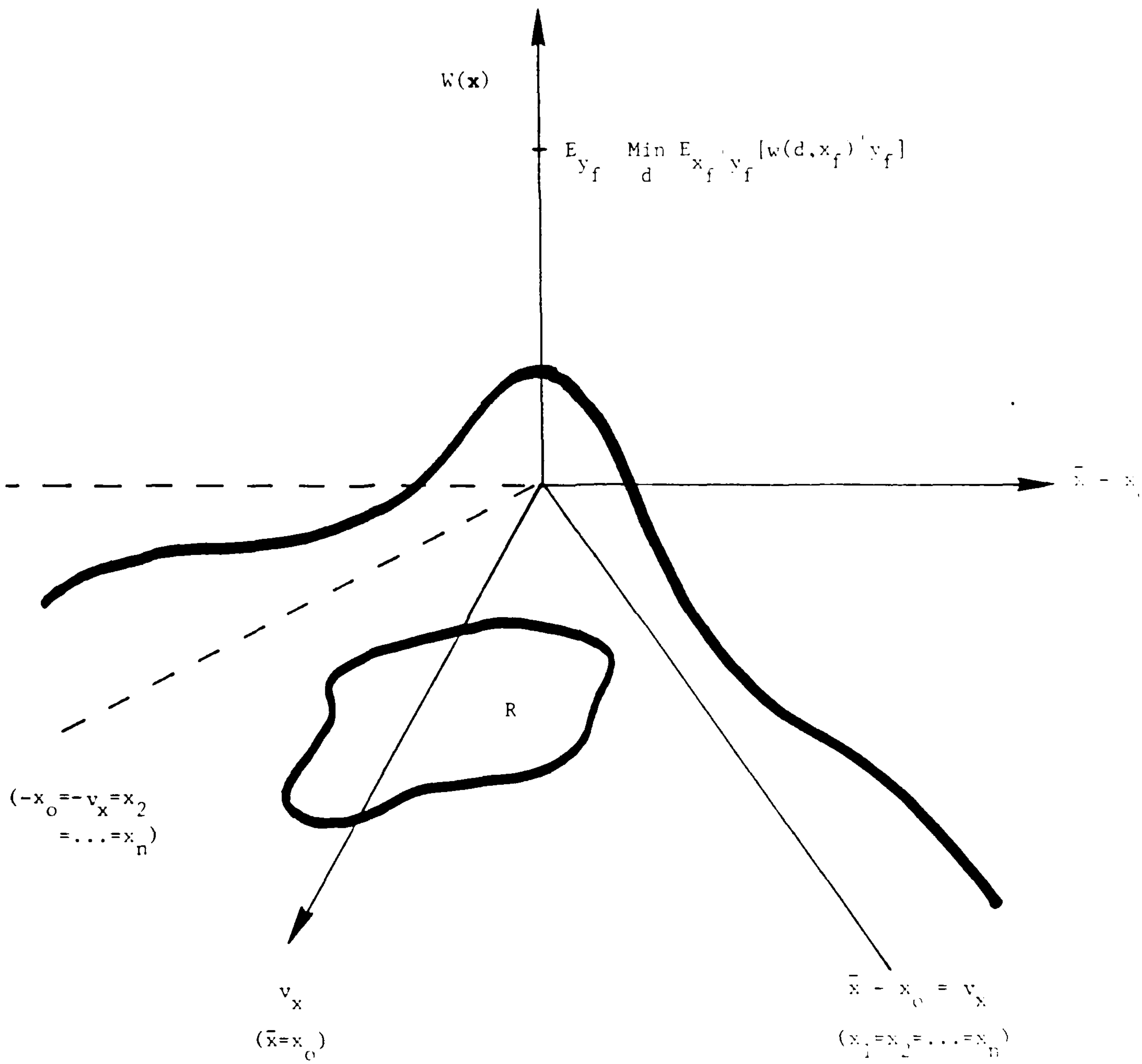
In this case

$$W(x) = E_{y|x}\, E_{y_f|y,x} \text{Var}(x_f|y_f,y,x)$$

$$= E_{y|x}\, E_{y_f|y,x}\, E_{x_f|y,x}(x_f^2|y_f,y,x) - E_{y|x}\, E_{y_f|y,x}\{E_{x_f|y,x}(x_f|y_f,y,x)\}^2.$$

Since x_f is independent of (x,y), we can explicitly evaluate the first term so that

$$W(x) = \sigma_o^2 + x_o^2 - E_{y|x}\, E_{y_f|y,x}\{E_{x_f|y,x}(x_f|y_f,y,x)\}^2.$$



$\sigma_b \neq 0$

($n \geq 2$ and fixed)

Figure 3.2

4. BIVARIATE NORMAL PRIOR FOR (α,β)

To calculate W(x) for a particular experimental design we need to assess a prior distribution for (α,β). Suppose $\alpha \perp \beta \perp \epsilon$ and α has a $N(a,\sigma_a^2)$ distribution while β has a $N(b,\sigma_b^2)$ distribution a priori. Table 4.1 gives the posterior bivariate normal parameters given the sufficient statistics n, $\bar{x} - x_o$, v_x, $z_1 = \sum_1^n e_i$ and $z_2 = \sum_1^n e_i(x_i - x_o)$. Note that σ_α^2, σ_β^2, and $\rho_{\alpha,\beta}$ do not depend on the observations, y, from the calibration experiment. The derivation of the posterior parameters in Table 4.1 is given in the appendix.

Our objective is to calculate W(x) for a given experimental design x. However, this is in general exceedingly difficult numerically. Hence we are also interested in bounds and efficient computational methods for special cases.

$$\mu_\alpha = a + \frac{(\Sigma e_i)[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2] - [\Sigma(x_i-x_o)][\Sigma e_i(x_i-x_o)]}{(n+\sigma^2/\sigma_a^2)[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2] - [\Sigma(x_i-x_o)]^2}$$

$$\mu_\beta = b + \frac{(n+\sigma^2/\sigma_a^2)[\Sigma e_i(x_i-x_o)] - [\Sigma(x_i-x_o)](\Sigma e_i)}{(n+\sigma^2/\sigma_a^2)[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2] - [\Sigma(x_i-x_o)]^2}$$

$$\sigma_\alpha^2 = \frac{\sigma^2[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2]}{(n+\sigma^2/\sigma_a^2)[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2] - [\Sigma(x_i-x_o)]^2}$$

$$\sigma_\beta^2 = \frac{\sigma^2(n + \sigma^2/\sigma_a^2)}{(n+\sigma^2/\sigma_a^2)[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2] - [\Sigma(x_i-x_o)]^2}$$

$$\rho_{\alpha\beta} = \frac{-\Sigma(x_i-x_o)}{\sqrt{(n+\sigma^2/\sigma_a^2)[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2]}}$$

$$\mathrm{cov}(\alpha,\beta) = \frac{-\sigma^2\Sigma(x_i-x_o)}{(n+\sigma^2/\sigma_a^2)[\Sigma(x_i-x_o)^2 + \sigma^2/\sigma_b^2] - [\Sigma(x_i-x_o)]^2}$$

where

$$e_i = y_i - a - b(x_i - x_o) .$$

TABLE 4.1. Parameters of the Posterior Distribution of (α,β) Given x and y

From the influence diagram, Figure 1.3, we see that at the time of decision, α and β are unknown. Hence we must first calculate the posterior distribution of α and β given n, $\bar{x} - x_o$, v_x, z_1 and z_2. The distribution of y_f given x_f, y and x is then

$$N[\mu_\alpha + \mu_\beta(x_f - x_o),\ s^2(x_f)]$$

where

$$s^2(x_f) = \sigma^2 + \sigma_\alpha{}^2 + \sigma_\beta{}^2(x_f - x_o)^2 + \\ + 2\mathrm{Cov}(\alpha,\beta)(x_f - x_o). \qquad (4.1)$$

Using Bayes' theorem

$$p(x_f|y_f,y,x) \;\; \alpha \;\; p(y_f|x_f,y,x)\; p(x_f)$$
$$\alpha \quad \exp\{-[y_f - \mu_\alpha - \mu_\beta(x_f - x_o)]^2/2s^2(x_f)\}\; p(x_f). \qquad (4.2)$$

Subtract μ_α from y_f and let

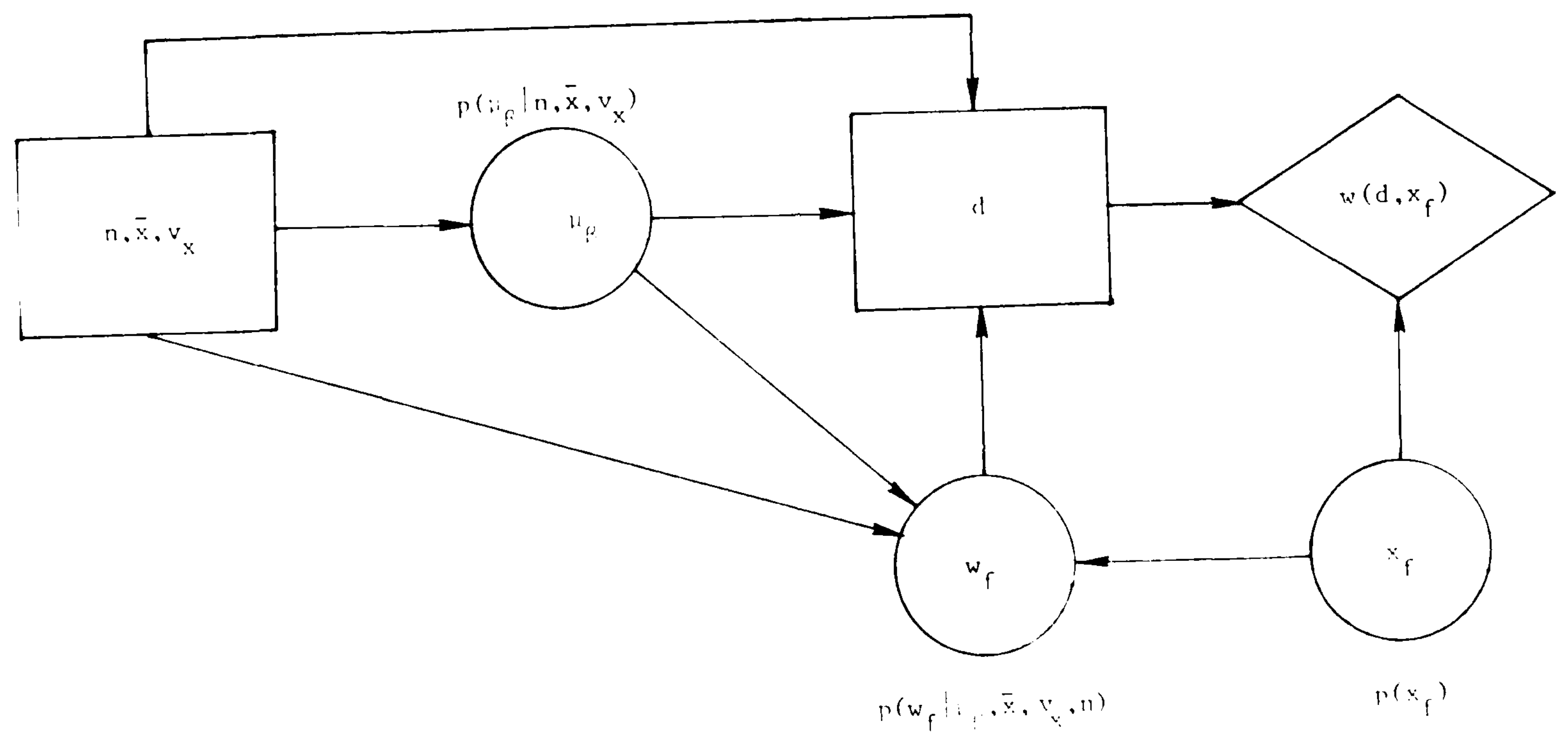
$$w_f = y_f - \mu_\alpha.$$

From (4.2), it is clear that

$$x_f \perp (y_f\ ,y)|\ w_f,\ \mu_\beta\ \ \ldots$$

i.e. w_f and μ_β are sufficient for x_f with respect to (y_f,y) where for convenience ... stands for all parameters which depend only on n, $\bar{x}$ and v_x. Since we consider n, $\bar{x}$ and v_x fixed and known in this section, we will omit these parameters in our conditioning statements. Also $\sigma_\alpha{}^2$, $\sigma_\beta{}^2$, and $\rho_{\alpha,\beta}$ depend only on n, $\bar{x}$, v_x and our bivariate normal prior parameter values. Hence these will also be omitted henceforth in our conditioning statements.

Based on sufficiency considerations for bivariate normal priors, the influence diagram in Figure 1.3 can be redrawn as in Figure 4.1. Note that whenever we needed to use Bayes' theorem (to achieve arrow reversals) it w



$$W(\mathbf{x}) = E_{\mu_\beta} E_{w_f \mid \mu_\beta} \operatorname*{Min}_{d} E_{x_f \mid w_f, \mu_\beta} [w(d, x_f) \mid w_f, \mu_\beta]$$

Figure 4.1

also helpful at that point to employ sufficiency considerations to reduce the parameter space.

Calculation of $W(x)$

(4.2) is the crux of our numerical difficulties since $p(x_f|w_f,\mu_\beta)$ is not normal even when x_f is $N(x_o,\sigma_o^2)$. Figures 4.2 and 4.3 are plots of $E(x_f|y_f)$ and $Var(x_f|y_f)$ versus y_f when the calibration experiment is not performed (i.e. n = 0). Were x_f and y_f jointly bivariate normal, $Var(x_f|y_f)$ would not depend on y_f as it obviously does in Figure 4.3.

Using Figure 4.1 we see that

$$W(x) = E_{\mu_\beta} E_{w_f|\mu_\beta} \underset{d}{\text{Min}}\, E_{x_f|w_f,\mu_\beta} [w(d,x_f)|w_f, \mu_\beta].$$

When $w(d,x_f) = (d - x_f)^2$ we have

$$W(x) = E_{\mu_\beta} E_{w_f|\mu_\beta} Var(x_f|w_f,\mu_\beta).$$

We can thus numerically calculate $W(x)$ using three nested subroutines for each x. The computational running time will be proportional to the product of the number of points used in each subroutine.

The Distribution of $w_f|x_f, \mu_\beta$

To calculate the posterior distribution of x_f given w_f and μ_β we need first to calculate the distribution of w_f given x_f and μ_β.

Theorem 4.1. $p(w_f|x_f,\mu_\beta)$ is $N[\mu_\beta(x_f - x_o), s^2(x_f)]$ where $s^2(x_f)$ is given by (4.1).

Proof. Clearly $E[w_f|x_f,\mu_\beta] = E_{\mu_\alpha} E[w_f|x_f,\mu_\alpha,\mu_\beta] = \mu_\beta(x_f - x_o)$. Since x_f is independent of (α,β) and y and (z_1,z_2) only appear in μ_α and μ_β

$$Var(w_f|x_f,\mu_\beta) =$$

$$E_{\mu_\alpha} [Var(w_f|x_f, \mu_\alpha, \mu_\beta)|x_f, \mu_\beta]$$

$$+ Var_{\mu_\alpha}[E(w_f|x_f, \mu_\alpha, \mu_\beta)|x_f, \mu_\beta].$$

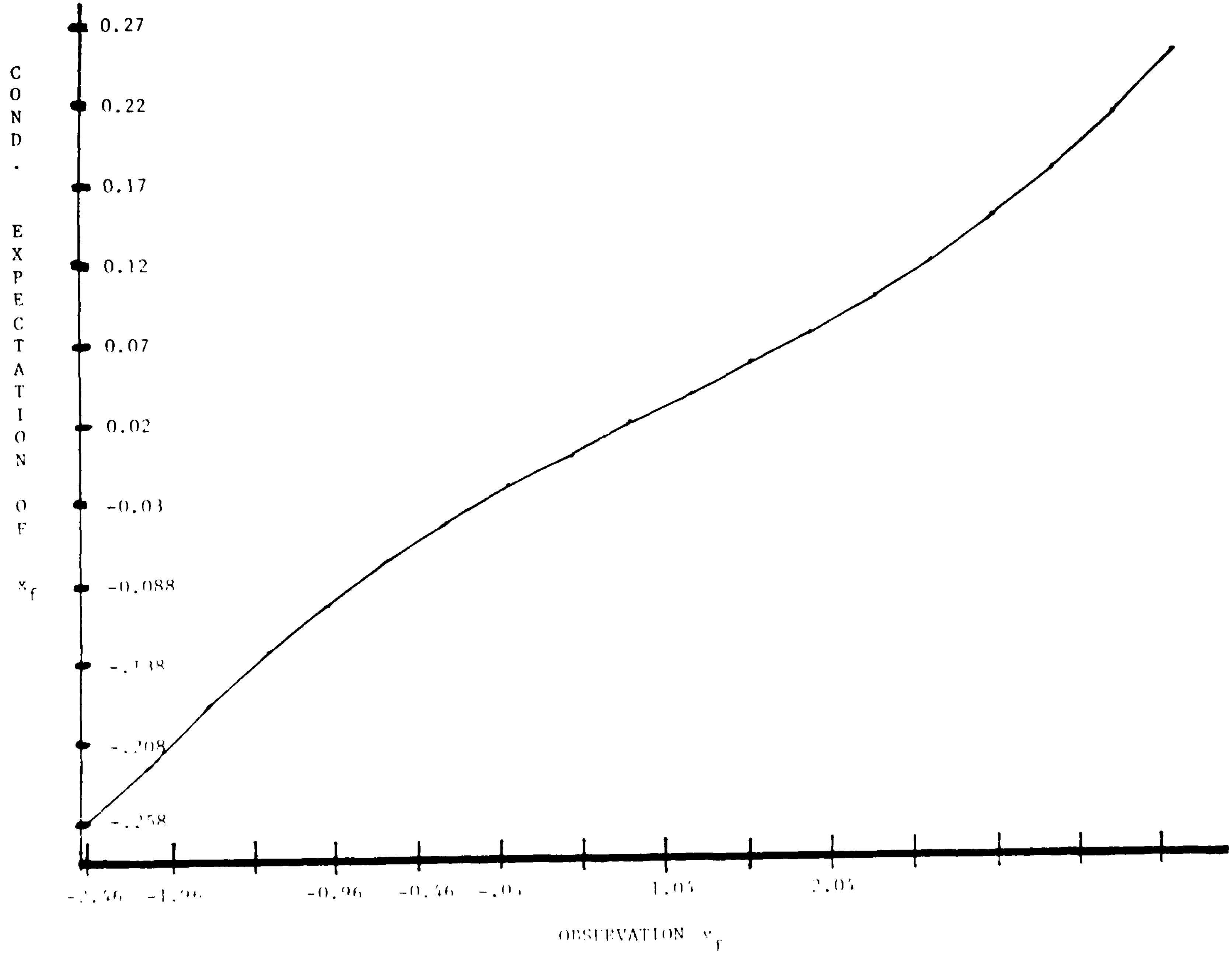
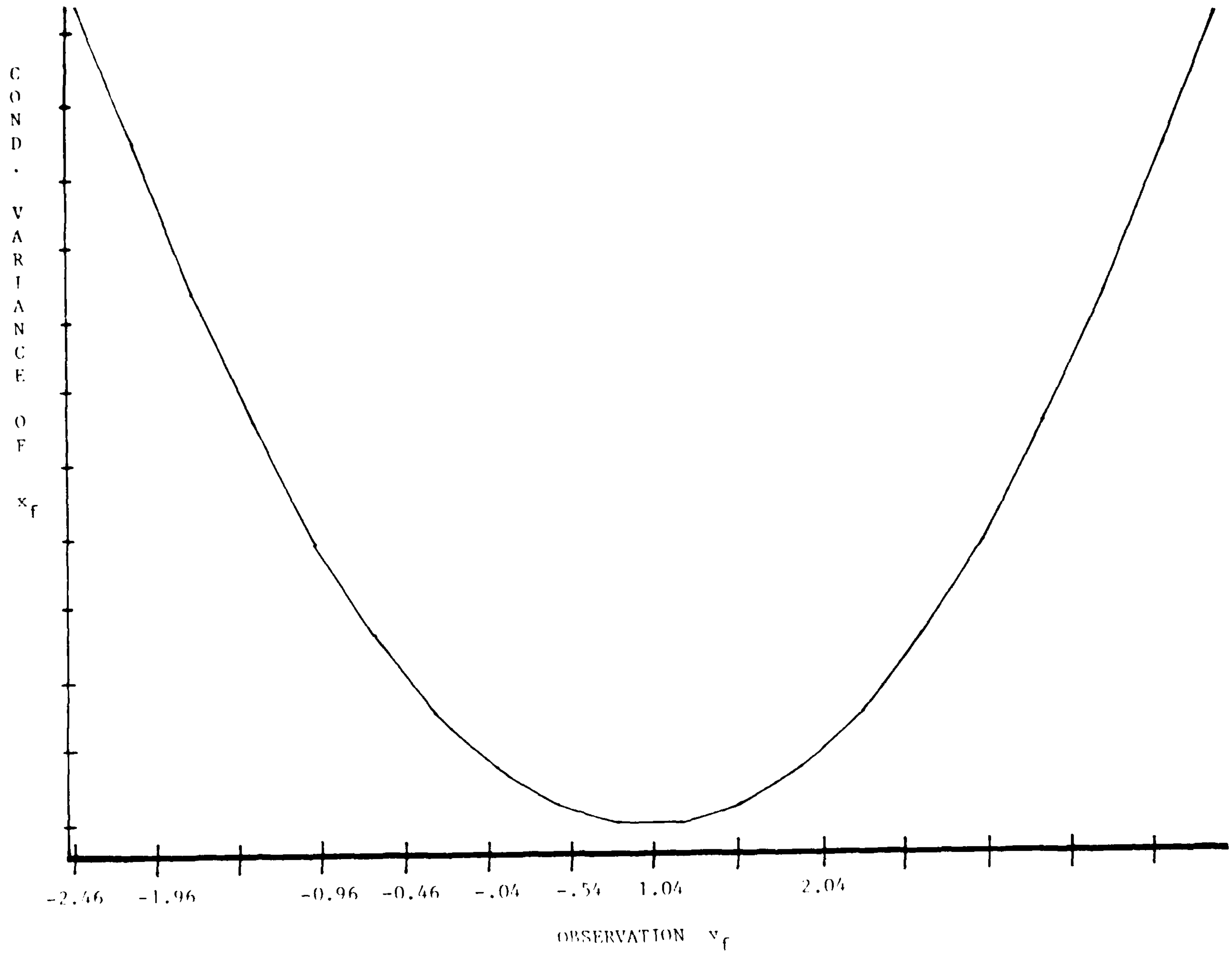

COND. EXPECTATION OF x_f
0.27
0.22
0.17
0.12
0.07
0.02
-0.03
-0.088
OBSERVATION y_f

Figure 5.2



MODEL: $E(v_f|x_f) = 1 + x_f$

$\sigma_a = 1$ $\sigma_b = 3$ Figure 4.3

From (4.1) we see that the first term is the same as $s^2(x_f)$ which is constant in (z_1, z_2) while the second term is 0. QED

5. NUMERICAL CALCULATION OF W(x) WHEN (α, β) IS BIVARIATE NORMAL

The Case When β is Known and $w(d, x_f) = (d - x_f)^2$

As we noted in Section 3, when $\sigma_b = 0$ a priori, $\text{Var}(x_f | w_f, \mu_\beta)$ depends on the experimental design only through n.

Theorem 5.1. If $\sigma_b = 0$, x_f is $N(x_o, \sigma_o^2)$ and $w(d, x_f) = (d - x_f)^2$ then

$$W(x) = \{ b^2/[\sigma^2 + \sigma_\alpha^2] + 1/\sigma_o^2\}^{-1}$$

where $\sigma_\alpha^2 = (n/\sigma^2 + 1/\sigma_a^2)^{-1}$.

Proof. Since $p(w_f | x_f, \alpha, \beta = b)$ is $N[b(x_f - x_o), \sigma^2]$ the predictive density for w_f, $p(w_f | x_f, \mu_b = b)$, is

$$N[b(x_f - x_o), \sigma^2 + \sigma_\alpha^2].$$

If $p(x_f)$ is $N(x_o, \sigma_o^2)$ a priori, then by Bayes' theorem

$$p(x_f | w_f) \propto p(w_f | x_f)\, p(x_f)$$

$$\propto \exp\{-[w_f - \mu_\beta(x_f - x_o)]^2/2(\sigma^2 + \sigma_\alpha^2)\} \exp[-(x_f - x_o)^2/2\sigma_o^2].$$

Collecting terms in the exponents we find

$$\text{Var}(x_f | w_f) = [b^2/(\sigma^2 + \sigma_\alpha^2) + 1/\sigma_o^2]^{-1}$$

while

$$E(x_f | w_f) = \frac{\{[b^2/(\sigma^2 + \sigma_\alpha^2)][w_f/b] + x_o/\sigma_o^2\}}{[b^2/(\sigma^2 + \sigma_\alpha^2) + 1/\sigma_o^2]}.$$

Since $\text{Var}(x_f | w_f)$ does not depend on w_f in this case,

$$W(x) = \text{Var}(x_f | w_f).$$ QED

The Case When $x_1 = x_2 = \dots = x_n = x_o$ and $w(d, x_f) = (d - x_f)^2$

In this case we can numerically calculate W(x) using two nested subroutines. Because of the comparative ease of computation, this is almost as good as a closed form solution.

As we saw in Section 3, this choice of x will provide no information about β. Hence $\mu_\beta = b$ and

$$W(x) = E_{w_f} \mathrm{Var}(x_f | w_f) \quad .$$

Thus only two nested subroutines are required. In this case w_f given x_f and

$$\mu_\beta = b \text{ is } N[b(x_f - x_o),\ s^2(x_f)]$$

where
$$s^2(x_f) = \sigma^2 + \sigma_\alpha^{\ 2} + \sigma_b^{\ 2}(x_f - x_o)^2$$

and $\sigma_\alpha^{\ 2} = (n/\sigma^2 + 1/\sigma_a^{\ 2})^{-1}$.

The General Case

To numerically calculate W(x) using three nested subroutines we need the density of μ_β. From Table 4.1 we see that μ_β is a linear combination of z_1 and z_2. Since z_1 and z_2 are unconditionally bivariate normal it follows that μ_β is $N(b, \sigma_{\mu_\beta}^{\ 2})$ where $\sigma_{\mu_\beta}^{\ 2}$ depends on the covariance matrix of (z_1, z_2). It is easy to verify that z_1 is $N(0, \sigma_1^{\ 2})$ where

$$\sigma_1^{\ 2} = n^2\sigma_a^{\ 2} + [\sum_i^n (x_1 - x_o)]^2 \sigma_b^{\ 2} + n\sigma^2$$

while z_2 is $N(0, \sigma_2^{\ 2})$ where

$$\sigma_2^{\ 2} = [\sum_1^n (x_i - x_o)]^2 \sigma_a^{\ 2} + [\sum_1^n (x_i - x_o)^2]^2 \sigma_b^{\ 2} + \sigma^2 \sum_1^n (x_i - x_o)^2 \ .$$

Jointly z_1 and z_2 given x, σ, a, b, σ_a, and σ_b are bivariate normal with covariance

$$\sigma_{12} = \mathrm{Cov}(z_1, z_2 | \sigma, a, b, x)$$
$$= n[\sum_1^n (x_i - x_o)]\sigma_a^2 + [\sum_1^n (x_i - x_o)] \sum_1^n (x_i - x_o)^2 \sigma_b^2 + \sigma^2 \sum_1^n (x_i - x_o) \quad .$$

Using Table 4.1, let

$$\mu_\beta = b + c_1 z_1 + c_2 z_2$$

where

$$c_1 = -[\sum_1^n (x_i - x_o)]/D$$
$$c_2 = (n + \sigma^2/\sigma_a^2)/D$$

and

$$D = (n + \sigma^2/\sigma_a^2)[\sum_1^n (x_i - x_o)^2 + \sigma^2/\sigma_b^2] - [\sum_1^n (x_i - x_o)]^2 \quad .$$

It follows that μ_β is $N(b, \sigma_{\mu_\beta}^2)$ where

$$\sigma_{\mu_\beta}^2 = c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 + 2c_1 c_2 \sigma_{12} \quad .$$

REFERENCES

Aitchison, J. and I.R. Dunsmore (1975). *Statistical Prediction Analysis.* Cambridge Univ. Press, Cambridge, England. (Paperback edition 1980).

Barlow, R. E. (1987). Using Influence Diagrams. Report no. ORC 87-1, Operations Research Center, U.C. Berkeley, Berkeley, CA 94720.

Brown, P.J. (1982). Multivariate Calibration. (With discussion). *J. Roy. Statist. Soc.*, *B*, 44, 287-321.

Brown, P.J. and Rolf Sundberg (1985). Confidence and Conflict in Multivariate Calibration. Universitet Stockholm Research Report no. 140, August 1985. (Inst. for Forsakringsmatematik ach Matematisk Statistik).

Chaloner, K. (1984). Optimal Bayesian Experimental Design for Linear Models. *Annals of Statistics*, *12*, 283-300.

Hoadley, B. (1970). A Bayesian look at inverse linear regression. *J. Amer. Stat. Assoc.*, *65*, 356-369.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory* MIT Press, Cambridge, Mass.

Shachter, R. (1986). Evaluating influence diagrams. In: *Reliability and Quality Control*. A. P. Basu, ed. Elsevier Science Publishers (North Holland), pp. 321-344.

APPENDIX

DERIVATION OF THE POSTERIOR PARAMETERS IN TABLE 4.1

Suppose x and y have joint density

$$p(x,y) \;\alpha\; \exp[-(ax^2 + bx + cy^2 + dy + exy)/2]$$

where a, b, c, d and e are constants. Then it follows that the pair (x,y) has a bivariate normal distribution; i.e.,

$$p(x,y) = \frac{\exp\{-[(x-\mu_\alpha)^2/\sigma_\alpha^2 - 2\rho_{\alpha\beta}(x-\mu_\alpha)(y-\mu_\beta)/\sigma_\alpha\sigma_\beta + (y-\mu_\beta)^2/\sigma_\beta^2]/2(1-\rho^2)\}}{2\pi\,\sigma_\alpha\sigma_\beta\sqrt{(1-\rho_{\alpha\beta}^2)}}$$

By matching coefficients in corresponding terms in the exponents μ_α, μ_β, σ_α, σ_β, and $\rho_{\alpha\beta}$ can be expressed in terms of a, b, c, d, and e.

The coefficent of x^2 is $a = 1/[\sigma_\alpha^2(1 - \rho_{\alpha\beta}^2)]$.

The coefficent of y^2 is $c = 1/[\sigma_\beta^2(1 - \rho_{\alpha\beta}^2)]$.

The coefficient of xy is $e = -2\rho/[(1 - \rho_{\alpha\beta}^2)\sigma_\alpha\sigma_\beta]$.

Now $4\rho^2 = e^2/ac$ implies $\rho = -e/s\sqrt{ac}$ and

$$\sigma_\alpha^2 = 1/a(1 - \rho_{\alpha\beta}^2) = 4c/[4ac - e^2]$$
$$\sigma_\beta^2 = 1/c(1 - \rho_{\alpha\beta}^2) = 4a/[4ac - e^2]$$

From the coefficient of x we have

$$b = [-2\mu_\alpha/\sigma_\alpha^2 + 2\rho_{\alpha\beta}\,\mu_\beta/\sigma_\alpha\sigma_\beta]/(1 - \rho_{\alpha\beta}^2)$$

while from the coefficient of y we have

$$d = [-2\mu_\beta/\sigma_\beta^2 + 2\rho_{\alpha\beta}\,\mu_\alpha/\sigma_\alpha\sigma_\beta]/(1 - \rho_{\alpha\beta}^2)$$

By taking σ_α times the first and $\rho_{\alpha\beta}\ \sigma_\beta$ times the second equation

$$b\sigma_\alpha + d\rho_{\alpha\beta}\sigma_\beta = -2\mu_\alpha/\sigma_\beta$$

so that

$$\mu_\alpha = -[b\sigma_\alpha{}^2 + d\rho_{\alpha\beta}\ \sigma_\alpha\sigma_\beta]/2$$
$$= -[b4c/(4ac - e^2) - de4\sqrt{ac}/2\sqrt{ac}(4ac - e^2)]/2$$
$$= [de - 2bc]/[4ac - e^2]$$

and also

$$\mu_\beta = [be - 2ad]/[4ac - e^2]$$

etc.

END

3-87

DTIC